Available online at www.sciencedirect.com
ScienceDirect
Hormones and Behavior 53 (2008) 177–184
Hormones and Behavior
www.elsevier.com/locate/yhbeh

Blocks and bodies: Sex differences in a novel version of the Mental Rotations Test

Gerianne M. Alexander *, Milagros Evardone

Department of Psychology, Texas A&M University, TAMU-4235, College Station, Texas 77845, USA

Received 25 June 2007; revised 17 September 2007; accepted 18 September 2007
Available online 29 September 2007

Abstract

A novel version of the Mental Rotations Test (MRT) that alternated the standard block figures with three-dimensional human figures was administered to 99 men and 129 women. Women and men differed predictably in their retrospective reports of childhood play and digit ratios, a putative measure of prenatal androgen action. Compared to the block figure items, human figure items on the modified MRT were associated with an improvement in performance in both sexes. However, consistent with the study hypothesis, the enhancing effect of the human figure condition on performance as measured by conventional scores was smaller in men compared to women and not at all evident in men when performance was measured by ratio scores. A closer inspection of the human figures effects on test scores showed performance in women improved for both male and female figure items. In contrast, relative to scores on block figure items, performance in men improved when stimuli were male figures but did not improve when stimuli were female figures. These results add to the evidence that the magnitude of sex differences in scores on the MRT may vary according to the test content and item properties. The findings suggest that online measures of cognitive processing in response to different classes of test stimuli (e.g., animate vs. inanimate objects, self-relevant vs. neutral stimuli) may prove useful in research aimed at understanding the hormonal and social factors contributing to the sex difference in performance on the MRT.

Keywords: Sex differences; Mental rotation; Human figures; Digit ratios

Introduction

Converging evidence from research on human and nonhuman animal species suggests that behaviors showing large sex differences are sensitive to hormonal influences (Alexander and Peterson, 2001; Cohen-Bendahan et al., 2005; Collaer and Hines, 1995). Therefore, research on human sex differences and their association to hormonal factors has often included the Mental Rotations Test (MRT) (Vandenberg and Kuse, 1978), a task that shows a robust male advantage in test scores (Voyer et al., 1995). The MRT measures how well individuals match objects that differ in their spatial orientation and, like all matching tasks, this multi-step cognitive process includes perceptual and decision making components (Parsons, 2003). Additionally, findings that reaction times and error rates increase with increasing angular differences between objects have indicated that comparisons are made after mentally rotating internal representations of objects to the same orientations (Shepard and Metzler, 1971). Results from research measuring brain activation (Hugdahl et al., 2006; Jordan et al., 2002) and patterns of visual attention (Alexander and Son, 2007) during task performance have suggested that this process differs between men and women, such that men may generally manipulate holistic representations (i.e., a mental image of the object is constructed and then rotated), whereas women may generally manipulate component part representations (i.e., they may focus on comparing object parts). Holistic representations correspond to a single unit that can be manipulated in the mind. However, piecemeal representations are perceived relations among multiple units or parts that must be preserved during spatial transformations. Therefore, compared to holistic representations, piecemeal representations are more difficult to sustain as angular differences between objects increase (for a discussion, see Amorim et al., 2006), consistent with the higher MRT error rates in women compared to men.

* Corresponding author.
E-mail address: gma@psyc.tamu.edu (G.M. Alexander).

0018-506X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.yhbeh.2007.09.014

Interestingly, the type of stimuli that must be represented mentally in these tasks appears to influence the variable magnitude of sex differences observed across measures (Voyer et al., 1995). For example, the MRT that uses three-dimensional block figures shows a large sex difference (Voyer et al., 1995), as does a similar test using two-dimensional abstract shapes (Collins and Kimura, 1997). In contrast, other tests such as those depicting two-dimensional panda bears (Grimshaw et al., 1995) or two-dimensional facial profiles (Merriman et al., 1985) do not. Tasks using abstract shapes or figures as test stimuli may yield large sex differences because, compared to meaningful stimuli, constructing and maintaining mental representations of abstract shapes are a more difficult task. It may also be significant in understanding the large sex differences in error rates on the MRT that experience or familiarity with stimuli increases the likelihood that holistic representations of such stimuli are constructed (Bethell-Fox and Shepard, 1988). As male-typical play includes stronger preferences for blocks and construction toys (Alexander and Hines, 1994), we hypothesized that the MRT may afford an advantage to men because male-typical play provides greater familiarity or comfort with figures constructed by cubes or blocks and so promotes the use of holistic strategies that facilitate task performance.

This proposed relationship between play experiences and scores on the MRT differs from the suggestion that the boys' higher level of experience with object manipulation, movement and construction (Sherman, 1967; Tracy, 1987) supports sex-linked cognitive strategies or brain processes that explain the persistent male advantage in spatial ability (Linn and Petersen, 1974; Maccoby and Jacklin, 1974; Voyer et al., 1995). The general understanding of that association between play and spatial ability is that the high spatial content of male-typical play facilitates the development of the general ability to perform spatial transformations (Voyer et al., 2000). However, both male-typical play and female-typical play include spatial manipulation of objects and the relevant spatial content of female-typical play styles (e.g., dressing dolls) is supported by findings that damage to brain parietal areas results in dressing difficulties as a function of deficits in mental rotation ability (Fitzgerald et al., 2002; Yamazaki et al., 2001). Thus, we reasoned that male-typical play may enhance the mental rotation of replicas of inanimate objects such as vehicles and blocks, whereas female-typical toy play, such as dressing dolls, may enhance the mental rotation of animate forms or body parts.

The goal of the present investigation was to examine sex differences in scores on a novel test of mental rotation ability using three-dimensional human figures. Importantly, recent researchers (Amorim et al., 2006) have established that matching three-dimensional human figures with different spatial configurations involves mental rotation, as indicated by an increase in reaction times and error rates with angular disparity between the human figures. As well-established body schemas in adults serve as a strong reference that promotes holistic mental rotation strategies (Amorim et al., 2006), we expected that scores in both sexes may be higher in the test of mental rotation ability using human figures relative to the test using block figures. However, in view of the evidence that males appear more likely than women to employ a holistic strategy in the typical MRT using block figures (Alexander and Son, 2007; Hugdahl et al., 2006), we hypothesized that the effect of human figures on scores would be larger in women compared to men.

Methods

Participants

Ninety-nine men (mean age = 19.67 years, SD = 1.87) and 129 women (mean age = 19.69 years, SD = 1.42) between 18 and 25 years of age were recruited from Introductory Psychology courses at Texas A&M University. Participants were tested during the months of October and November in eight separate group sessions of up to 30 people. All women and men provided informed consent and received partial credit towards a course requirement.

Measures

Subject characteristics

The Extended Range Vocabulary Test (Ekstrom et al., 1976) was included as a control measure of general cognitive ability that does not show a sex difference. The measure loads strongly on factors identified as indicating crystallized intelligence (e.g., Rossman and Horn, 1972), has good reliability (0.76 = 0.89) (Ekstrom et al., 1976), and is used widely in research on sex differences in cognitive abilities to support the specificity of any sex differences in task performance (e.g., Alexander et al., 1998; Chipman and Hampson, 2006; Choi et al., 2006; Miles et al., 2006).

The ratio of the lengths of the second and fourth digits of the right hand (2D:4D) and the Pre-School Activities Inventory (PSAI; Golombok and Rust, 1993) were included as measures of characteristics that typically differ between the sexes to support the representational nature of the sample. The 2D:4D ratio is a sexually dimorphic trait that some evidence suggests is a direct correlate of prenatal sex steroid levels (Lutchmaya et al., 2004; Manning et al., 1998), such that smaller ratios are correlated with higher androgen levels. However, a recent proposal is that digit ratios may be better described as a measure of perinatal androgen action (McIntyre, 2006), consistent with findings that smaller digit ratios are associated with androgen receptor alleles showing fewer terminal domain CAG repeats, a marker of greater androgen sensitivity (Manning et al., 2003). In this research, the 2D:4D ratio was calculated by obtaining a digital photo scan of the participant's right hand. Color images of hands were later used to measure the distance in millimeters from the basal crease to the tip of the second and fourth fingers with digital vernier calipers. Two independent judges coded finger lengths for each hand copy. Consistent with the findings of previous research using this method of assessment (Alexander, 2006; McIntyre, 2006), measurements averaged across the two judges showed excellent inter-rater reliability (rs > .95).

The PSAI is a 24-item questionnaire measuring sex-linked childhood play preferences, including toys, activities, and characteristics. The PSAI shows that moderate test–retest reliability ($r = .62 = r = .66$) and PSAI scores in children correlate moderately with teacher ratings of behavior (boys = $r = .37$, girls = $r = .48$) (Golombok and Rust, 1993). When used with adults (Alexander, 2006; Hines et al., 2004), the sex difference in recalled childhood preferences is very large ($d = 2.65–3.25$).

Dancing aggies: a modified Mental Rotations Test

The redrawn Vandenberg and Kuse (1978) Mental Rotations Test (MRT-1; Peters et al., 1995) consists of 24 items depicting a target and four alternatives (two correct and two distracters). The modified task consisted of 12 original test items alternating with 12 test items depicting human figures, such that items 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 were the modified human figure items and items 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24 corresponded to the original task items of the same number. The 12 human figure items, constructed using Autodesk Maya 6.5 software, depicted three-dimensional men (items 1, 5, 9, 13, 17, and 21) and women (items 3, 7, 11, 15, 19, and 23) dressed in identical T-shirts and pants. Items from the original task were used as templates for the spatial orientation of the figure in three-dimensional space and for the serial positioning of correct items and distracters. Like the MRT block figure items, human figure items included occlusions of some features (see Fig. 1), a test item characteristic

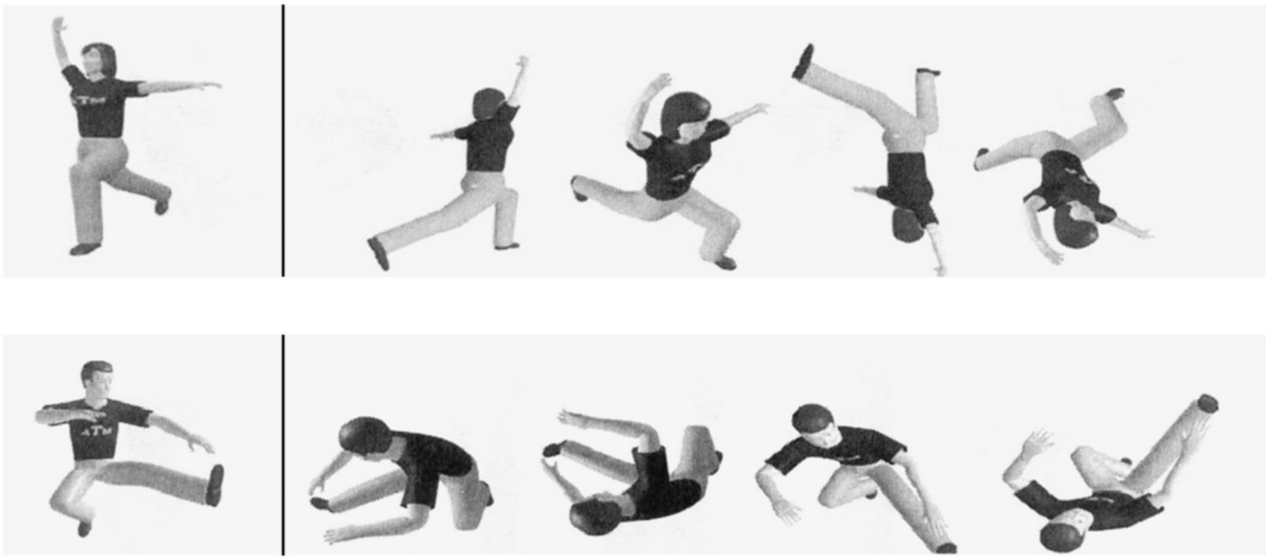

Fig. 1. Examples of the human figure items used in the modified task.

of the MRT that yields the largest sex differences in number correct (Voyer and Hou, 2006). For all items, participants were asked to choose the two alternatives that portrayed a rotated representation of the stimulus object. As in earlier research (Peters et al., 1995), they were given 3 min to complete the first 12 items and 3 min for the remaining 12 items, with a 2-minute break in between.

MRT scoring method

Following from recent research (e.g., Geiser et al., 2006), two measures of performance, *conventional scores* and *ratio scores*, were considered in the analysis of sex differences. Conventional scores are the total number of responses identifying both correct items, a widely used scoring method that produces the largest sex difference in performance (Voyer et al., 1995). Accordingly, the maximum conventional score in this research for both block figure items and human figure items was 12. Ratio scores correct for individual differences in the number of items completed by dividing the total number items correct by the total number of items attempted.

Procedure

All participants completed the measures in the following order: (1) Pre-School Activities Inventory, (2) Vocabulary, and (3) the modified version of Mental Rotations Test (MRT-1) (Peters et al., 1995; Vandenberg and Kuse, 1978). Instructions for each measure were read aloud and participants were provided the same amount of time to complete each measure individually. If finished before others, participants were instructed not to proceed to the next task. Once all participants were finished with one task, they were able to proceed as a group to the next task. At the end of the 30-minute session, participants were verbally debriefed and when possible photocopies of participants' right hands were obtained.[1]

Results

Subject characteristics

Scores on the vocabulary test were similar in women ($M=24.24\pm7.24$) and men ($M=24.22\pm6.96$), $t(226)<1.0$, $d=0.002$ (Cohen, 1977). Compared to women, men reported more male-typical behavior on the PSAI ($M=75.99\pm9.50$ vs. $M=31.79\pm14.38$), $t(226)=26.45$, $p<.001$, $d=3.62$. Digit ratio measurements collected from 59 men and 77 women showed smaller (i.e., more male-typical) second to fourth digit ratios in men ($M=0.95\pm0.02$) than in women ($M=0.98\pm.03$), $t(135)=5.37$, $p<.001$, $d=0.94$. Thus, scores on the general measure of cognitive ability were comparable in women and men who also differed predictably in the two measures of early sex-linked characteristics.

Mental rotation

Only 12% of women and 12% of men completed all 24 items of the modified MRT. Table 1 summarizes the completion rates for women and men within each condition of the task. Completion rates were examined using ANOVA for repeated measures with sex (male vs. females) as a grouping factor and figure type (block figure vs. human figure) as the within-subject factor. This analysis showed main effects of sex, $F(1,226)=13.56$, $p<.001$, and figure type, $F(1,226)=49.99$, $p<.001$, and a sex by figure type interaction, $F(1,226)=7.95$, $p<.01$. The main effect of sex occurred because men completed significantly more test items than did women ($M=19.42\pm3.88$, range: 7–24 items vs. $M=17.36\pm4.56$, range: 5–24 items, $d=0.49$). The

Table 1
Scores (means and standard deviations) for women and men on the block figure items and human figure items of the modified Mental Rotations Test

	Women	Men	Effect size (d)
Block figures			
Number completed[a]	8.08 (2.82)	9.44 (1.91)	0.56
Conventional score[a]	3.85 (2.34)	6.74 (2.96)	1.08
Ratio score	0.48 (0.27)	0.71 (0.27)	0.85
Human Figures			
Number completed[a]	9.28 (2.27)	9.98 (2.10)	0.32
Conventional score[a]	5.95 (2.86)	7.41 (2.51)	0.54
Ratio score	0.64 (0.25)	0.74 (0.19)	0.45

[a] Maximum score = 12.

[1] Testing occurred in classrooms between regularly scheduled lecture times. On two occasions, the classroom did not remain unoccupied long enough to collect photocopies of participants' hands.

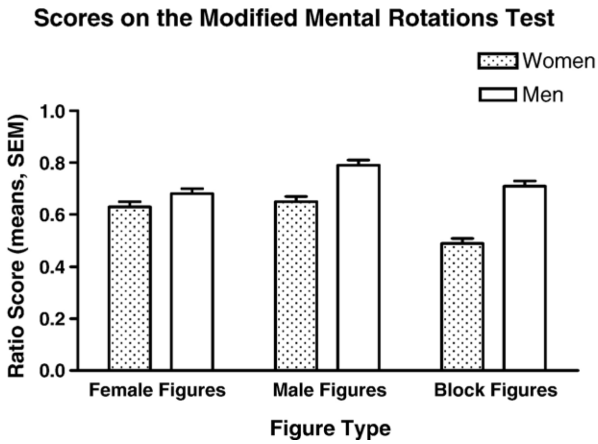

Fig. 2. Ratio scores for female human figure items, male figure items, and block figure items in women and men illustrating the improved performance of women under both human figure conditions ($p<.05$) and the improved performance of men under the male human figure condition ($p<.05$).

main effect of figure type occurred because completions were generally higher for the human figure condition compared to the block condition ($d=0.36$). Tests of the simple effects of figure type within each level of sex showed a higher completion rate of human figure items in men, $F(1,226)=8.00$, $p<.01$, and in women, $F(1,226)=56.09$, $p<.001$, although the effect size of the difference in completion rates was moderate in women ($d=0.48$) and small in men ($d=0.26$). The sex differences in completion rates supported the additional analysis of ratio scores.

Table 1 summarizes the group means and the magnitude of the sex differences in the two performance measures (conventional scores and ratio scores). The correlation between scores for block figure items and scores for human figure items was significant in men ($r(98)=0.52$ for conventional scores and $r(98)=0.43$ for ratios scores, ps$<.001$) and in women ($r(128)=0.35$ for conventional scores and $r(128)=0.41$ for ratios scores, $p<.001$). There was no significant difference between the correlation coefficients in women and men for either measure of performance (Fisher $z=1.56$, $p=.11$ for conventional scores and Fisher $z=0.18$, *ns*, for ratio scores).

Conventional scores and ratio scores were analyzed using separate ANOVAs for repeated measures with sex (male vs. females) as a grouping factor and figure type (block figure vs. human figure) as the within-subject factor. The analysis of conventional scores showed main effects of sex, $F(1,226)=53.62$, $p<.001$, and figure type, $F(1,226)=50.52$, $p<.001$, and a sex by figure type interaction, $F(1,226)=13.07$, $p<.001$. The main effect of sex occurred because average scores across the two conditions were generally higher in men compared to women ($d=0.97$). The main effect of figure type occurred because scores were generally higher for the human figure condition ($d=0.50$). Tests of the simple effects of figure type within each level of sex showed significant effects of figure type in men, $F(1,226)=5.38$, $p<.05$, and in women, $F(1,226)=66.57$, $p<.001$, although the effect size was large in women ($d=0.80$) and small in men ($d=0.25$).

The analysis of ratio scores also showed main effects of sex, $F(1,226)=32.51$, $p<.001$, and figure type, $F(1,226)=20.06$, $p<.001$, and a sex by figure type interaction, $F(1,226)=9.06$, $p<.01$. The main effect of sex occurred because average ratio scores across the two conditions were generally higher in men compared to women ($d=0.76$). The main effect of figure type occurred because ratio scores were generally higher for the human figure condition compared to the block condition ($d=0.34$). Tests of the simple effects of figure type within each level of sex showed significant effects of figure type on performance in women, $F(1,226)=32.02$, $p<.000$, $d=0.58$, but not in men, $F(1,226)=0.96$, *ns*, $d=0.08$.

Human figure sex

To explore the possible differential effects of human figure sex on performance in women and men relative to their performance on the block figure items, ratio scores were analyzed again using ANOVA for repeated measures with sex (men vs. women) as a grouping factor and ratio scores (block score, female figure score, male figure score) as the within-subject factor. That analysis showed a significant sex by figure type interaction, $F(2,224)=8.98$, $p<.001$. Simple main effects of figure type on performance were significant in men, $F(2,224)=8.44$, $p<.001$, and in women, $F(2,224)=16.37$, $p<.001$. Ratio scores in women were similar for both male and female figures, $d=0.10$, and pairwise comparisons of the means for block, female, and male items showed women's scores for both male and female human figure items were significantly higher than their block figure item scores (ps$<.05$). In contrast, ratio scores in men were higher for male figure items compared to female figures ($d=0.45$). Furthermore, pairwise comparisons of the means for the three figure types showed men's scores for male human figure items were higher than their scores for block items ($p<.05$), but men's scores for block items and female figure items did not differ ($p=.30$) (Fig. 2). Interestingly, a similar analysis using the smaller number of women and men completing all items of the mental rotations replicated the sex by figure type interaction, $F(2,25)=4.66$, $p<.01$, and the

Table 2
Within-sex correlations between task performance and the behavioral measures

	Vocabulary	PSAI	Digit ratios
Men (*n=99*)			
Block figures			
Conventional scores	0.32**	0.12	0.18
Ratio scores	0.19	0.04	0.02
Human figures			
Conventional scores	0.27**	0.09	0.32*
Ratio scores	0.15	0.08	0.19
Women (*n=129*)			
Block score			
Conventional scores	0.03	−0.10	0.20
Ratio scores	−0.07	−0.01	0.25*
Human figures			
Conventional scores	0.19*	0.03	0.07
Ratio scores	0.05	0.06	0.15

* denotes significance at $p<.05$, ** at $p<.01$.

direction and significance of the group differences in performance.

Finally, Table 2 summarizes the within-sex correlations between performance scores and the measures of subject characteristics. There was a significant difference between the two correlation coefficients measuring the association between vocabulary and block item scores in women and men (conventional scores, Fisher $z=2.23$, $p<.05$; ratio scores, Fisher $z=1.95$, $p=.05$). No other correlation coefficients differed significantly between the sexes. Not shown in Table 2 was the finding that the correlation between digit ratios and scores on PSAI in both sexes was small and insignificant, $r<.10$. Furthermore, the relationship between PSAI scores and performance on the human figures test was not altered in either sex when only items using male figures were used to calculate the measure of performance (all r's$<.10$).

Discussion

We theorized that the large sex difference in error rates on the widely used Mental Rotations Test (Voyer et al., 1995) may result in part because the test stimuli used in that task afford an advantage to men independent of levels of spatial ability. Therefore, in this research a novel version of the task was constructed by substituting three-dimensional human figures for the three-dimensional block figures in half of the test items. Women and men completing this task did not differ on a measure of general cognitive ability (i.e., vocabulary) but did differ predictably on measures of sex-linked characteristics (childhood play and digit ratios). Also consistent with the general findings from previous research on sex-linked spatial abilities (Voyer et al., 1995), scores on the modified MRT were generally higher in men than in women. However, as predicted, women's test performance improved in the human figure condition, such that the magnitude of the sex difference in test scores was reduced by approximately half. At a general level, these results add to the evidence that the magnitude of sex differences in scores on the MRT may vary according to the test content and item properties (Voyer and Hou, 2006). In addition, understanding why these results occurred may increase understanding of the hormonal and social influences on sex-linked spatial ability, at least as measured by this widely used task.

Body schemas promote holistic representations of human figures (Amorim et al., 2006) that are thought to enhance mental rotation ability, and so we predicted that the human figure version of the MRT would reduce error rates in both women and men. Furthermore, as results from research measuring brain activation (Hugdahl et al., 2006; Jordan et al., 2002), patterns of visual attention (Alexander and Son, 2007), and patterns of solution probabilities (Geiser et al., 2006) suggest that women are less likely to apply holistic strategies when completing the MRT, we predicted that the promotion of holistic strategies by human figures would result in larger improvements in test scores in women compared to men. Consistent with these hypotheses, in both sexes, test scores derived from human figure items were generally higher than test scores derived from the block figure items. In addition, the enhancing effect of the human figure condition on performance as measured by conventional scores was smaller in men compared to women and not at all evident in men when performance was measured by ratio scores. A closer inspection of the human figures effects on test scores showed performance in women improved for both male and female figure items. In contrast, relative to scores on block figure items, performance in men improved when stimuli were male figures but did not improve when stimuli were female figures. The unexpected finding that performance in men varied according to the sex of the human figure is not predicted by the hypothesized increased use of holistic strategies for human figure items and, therefore, requires further explanation.

Women and men differ in their affective response to visual sexual stimuli, such that men show greater affective response to preferred sexual stimuli, as indicated by measures of physiological arousal (Chivers et al., 2004) and activation of brain areas for emotion (i.e., the amygdala) (Hamann et al., 2004). Although the human figures depicted in the modified MRT were clothed in a modest fashion, it may be that affective response of women and men in this research to task stimuli also differed, such that men showed greater affective response to female figures. If so, then emotional processes activated by human figures may have influenced test scores, a possibility consistent with animal research showing that the amygdala modulates brain systems for spatial abilities (Packard and Teather, 1998). However, as it is highly improbable that the majority of men in this research preferred same-sex partners, it is not obvious how a greater emotional response to female figures (presumably the preferred sexual stimulus) could account for the findings that men's test scores for female human figure items and block figure items (presumably a non-sexual stimulus) were similar and lower than their test score for male human figure items.

Several findings from research on human sex differences in social development may better explain the pattern of results for women and men. Men compared to women show stronger adherence to gender-linked roles, even in childhood (Ruble and Martin, 1998). One possibility is that men also develop body schemas that are more gender polarized with regard to sex-linked characteristics, such as the social conventions for hairstyle or the biologically determined secondary sex characteristics evident in the task stimuli. If so, then sex differences in the degree to which body schemas are self-referent (i.e., "like me") may contribute to differences in the efficiency of processing opposite sex human figures, such that men are more responsive to self-referent body stimuli. Some support for this hypothesis is provided by recent research on face processing demonstrating that men compared to women show increased brain activation in response to higher facial resemblance (Platek et al., 2005). At a more abstract level of cognitive representation, others have proposed that self representations characterized by connections with other people (i.e., relatedness to others) as opposed to independence or autonomy are more characteristic of women than of men, and so social stimuli may generally be more self-relevant to women than to men (Cross and Madson, 1997). Because self-relevant stimuli are associated with more efficient cognitive processing (e.g., Bargh, 1982), these possible sex differences in body schema or self representations may be factors contributing to the

present finding that opposite sex human figures afforded a task advantage to women but not to men.

Although this study was not designed to measure the underlying cognitive processes that may have contributed to the sex differences found in response to task stimuli, the general findings of this research are consistent with the hypothesized associations between sex-linked developmental factors and performance on the two versions of the MRT. For example, the suggestion that the results for the human figure items may have occurred because women develop flexible body schemas and representations of an interdependent self are consistent with the experiences afforded by female-preferred play styles that include playing with dolls or human figures of both sexes (e.g., Barbie, Ken). However, scores on the PSAI, a measure of gender-linked childhood activities and play, were unrelated to scores on the modified MRT. Thus, the novel proposal based on neuropsychological research (Fitzgerald et al., 2002; Yamazaki et al., 2001) that female-typical play may facilitate the mental rotation of animate forms was not supported in this research by the correlations between childhood play and performance. As other findings suggest that the contribution of male-typical toy play to spatial ability is relatively small (Baenninger and Newcombe, 1989; Voyer et al., 2000) and gender stereotypes may have biased recall of play styles in this research, our power to detect significant associations between the retrospective measure of sex-linked behavior and performance may be limited. Certainly, the sex difference that we observed in PSAI scores ($d>3.0$) is larger than that typically observed in children's play (Hines, 2006), suggesting that adults' questionnaire responses may have been biased by a general tendency to selectively recall gender-typical behavior (Martin, 1999). It may also be relevant in this regard that vocabulary, an approximate measure of crystallized intelligence (i.e., previously acquired knowledge) (Rossman and Horn, 1972), correlated in men with scores on both conditions of the modified MRT and in women with scores only on the human figure items. It is possible that this association between performance and a measure that loads heavily on factors described as previously acquired knowledge reflects the shared contribution of experiential factors to test scores. Female-typical experiential factors were hypothesized in this research to be less applicable to the mental rotation of block figure items, consistent with observed pattern of correlations. For all these reasons, a role for early experiences in the development of sex differences in performance on the modified MRT merits further investigation.

Finally, although significant associations between test scores and the ratio of the second to fourth digits in men have been reported (Kempel et al., 2005; Manning and Taylor, 2002; McFadden and Shubel, 2003), our findings add to the previous research showing no association between digit ratios indicating stronger prenatal androgen action (McIntyre, 2006) and MRT scores (Alexander and Son, 2007; Austin et al., 2002; Coolican and Peters, 2003; Poulin et al., 2003). Digit ratios indicating weaker androgen action in prenatal life were significant correlates of men's conventional scores for human figure items and women's ratio scores for block figure items. A similar positive association between digit ratios and MRT scores has been reported previously in women (Putz et al., 2004). However, such a relationship between prenatal hormones and behavior is not predicted by the general assumptions of hormonal theories of sex-linked behavior and so will require further replication to establish its reliability and theoretical implication. Importantly, test scores are endpoints of multiple processes that activate multiple brain systems (Parsons, 2003) and other aspects of task performance not dependent on verbal reports appear more responsive to hormonal factors, such as the identification or discrimination of relevant task features (Alexander and Son, 2007; Hooven et al., 2004). Therefore, it may be informative in future research on hormonal influences on mental rotation ability using human figures to include measures that are more sensitive to online visual and cognitive processes (e.g., visual attention to task stimuli, reaction times).

To our knowledge, this is the first investigation of sex differences in an analogue of the MRT using three-dimensional human figures. The modified MRT was administered in this research under timed conditions, as is typically the case in research on human sex differences (Voyer et al., 1995), and the modified MRT was also the last measure to be completed in a short battery of tasks. Timed conditions and the possibility that motivation dropped differentially in women and men across time are testing characteristics that may increase the magnitude of sex differences in MRT performance (Voyer, 1997), consistent with the present finding of a very large sex difference in conventional scores for the block figure items ($d>1.0$). If so, then the smaller sex difference on human figure items documented in this research may also be an overestimation of group differences, a possibility that would strengthen our interpretation of the present results and so would be important to determine in future research using the task. In addition, only one version of the modified task was constructed. Although the modified task preserved the item order of the original MRT, the presentation of human figures in the first item of the test is a limitation of the research design that may have contributed to the higher completion rates for human figure items compared to block figure items ($d=0.48$ in women and $d=0.26$ in men) and the large sex difference in test scores for block items. However, item order is unlikely to fully explain the between-sex and within-sex differences in the measures of error rates for human figure items, which included a measure correcting for the number of completed items and was also replicated in the analyses of women and men completing all test items. Nonetheless, recent evidence from research using multigroup latent class analysis of solution strategies for each item of the original MRT has indicated that strategies (holistic, analytic, or piecemeal) vary across the test items (Geiser et al., 2006). Therefore, future research examining the item properties of the MRT under both presentation conditions (block figures, human figures) may be valuable in better understanding the results that we report in this first investigation of the modified task. In summary, the present investigation indicates the potential usefulness of including multiple stimuli in research on sex differences in mental rotation. The results also suggest that a better understanding of the hormonal and social factors underlying the large sex difference in test scores on the MRT may emerge from a collaborative interface between hormone–behavior research and research on the cognitive processes

underlying mental rotation and spatial transformations of different classes of stimuli (e.g., animate vs. inanimate, self-referent vs. neutral).

Acknowledgments

This research was supported in part by National Institute of Mental Health grant MH071414 (GMA). The authors thank Brooke Beane for the construction of the task stimuli and Mark G. Packard and two anonymous reviewers for helpful comments on an earlier version of the manuscript.

References

Alexander, G.M., 2006. Associations among gender-linked toy preferences, digit ratio and spatial ability: evidence from eye-tracking analysis. Arch. Sex. Behav. 35, 699–709.

Alexander, G.M., Hines, M., 1994. Gender labels and play styles: their relative contribution to children's selection of playmates. Child Dev. 65, 869–879.

Alexander, G.M., Peterson, B.S., 2001. Sex steroids and human behavior: implications for developmental psychopathology. CNS Spectr. 6, 75–88.

Alexander, G.M., Son, T., 2007. Androgens and eye movements in women and men during a test of mental rotation ability. Horm. Behav. 52, 197–204.

Alexander, G.M., Swerdloff, R.S., Wang, C., Davidson, T., McDonald, V., Steiner, B., Hines, M., 1998. Androgen–behavior correlations in hypogonadal men and eugonadal men. II. Cognitive abilities. Horm. Behav. 33, 85–94.

Amorim, M., Isableu, B., Jarraya, M., 2006. Embodied spatial transformations: "body analogy" for the mental rotation of objects. J. Exp. Psychol. Gen. 135, 327–347.

Austin, E.J., Manning, J.T., McInroy, K., Mathews, E., 2002. A preliminary investigation of the associations between personality, cognitive ability and digit ratio. Pers. Individ. Differ. 33, 1115–1124.

Baenninger, M., Newcombe, N., 1989. The role of experience in spatial test performance: a meta-analysis. Sex Roles 20, 327–344.

Bargh, J.A., 1982. Attention and automaticity in the processing of self-relevant information. J. Pers. Soc. Psychol. 43, 425–436.

Bethell-Fox, C.E., Shepard, D.M., 1988. Mental rotation: effects of stimulus complexity and familiarity. J. Exp. Psychol. Hum. Percept. Perform. 14, 12–23.

Chipman, K., Hampson, E., 2006. A female advantage in the serial production of non-representational learned gestures. Neuropsychologia 44, 2315–2329.

Chivers, M.L., Rieger, G., Latty, E., Bailey, M., 2004. A sex difference in the specificity of sexual arousal. Psychol. Sci. 15, 736–744.

Choi, J., McKillip, E., Ward, M., L'Hirondelle, N., 2006. Sex-specific relationships between route-learning strategies and abilities in a large-scale environment. Environ. Behav. 38, 791–801.

Cohen-Bendahan, C.C.C., van de Beek, C., Berenbaum, S.A., 2005. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. Neurosci. Biobehav. Rev. 29, 353–384.

Cohen, J., 1977. Statistical Analysis for the Behavioral Sciences. Academic Press, New York.

Collaer, M.L., Hines, M., 1995. Human behavioral sex differences: a role for gonadal hormones during early development? Psychol. Bull. 118, 55–107.

Collins, D.W., Kimura, D., 1997. A large sex difference on a two-dimensional mental rotation task. Behav. Neurosci. 111, 845–849.

Coolican, J., Peters, M., 2003. Sexual dimorphism in the 2D/4D ratio and its relation to mental rotation performance. Evol. Hum. Behav. 24, 179–183.

Cross, S.E., Madson, L., 1997. Models of the self: Self-construals and gender. Psychol. Bull. 122, 5–37.

Ekstrom, R.B., French, J.W., Harman, H.H., 1976. Kit of Factor Referenced Cognitive Tests. Educational Testing Service, Princeton, NJ.

Fitzgerald, L.K., McKelvey, J.R., Szeligo, F., 2002. Mechanisms of dressing apraxia: a case study. Neuropsychiatry Neuropsychol. Behav. Neurol. 15, 148–155.

Geiser, C., Lehmann, W., Eid, M., 2006. Separating "rotators" from "nonrotators" in the Mental Rotations Test: a multigroup latent class analysis. Multivariate Behav. Res. 41, 261–293.

Golombok, S., Rust, J., 1993. The Pre-School Activities Inventory: a standardized assessment of gender role in children. Psychol. Assess. 5, 131–136.

Grimshaw, G.M., Sitarenios, G., Finegan, J.K., 1995. Mental rotation at 7 years: relations with prenatal testosterone levels and spatial play experiences. Brain Cogn. 29, 85–100.

Hamann, S., Herman, R.A., Nolan, C.L., Wallen, K., 2004. Men and women differ in amygdala response to visual sexual stimuli. Nat. Neurosci. 7, 411–416.

Hines, M., 2006. Prenatal testosterone and gender-related behaviour. Eur. J. Endocrinol. 155, S115–S121.

Hines, M., Brook, C., Conway, G.S., 2004. Androgen and psychosexual development: core gender identity, sexual orientation, and recalled childhood gender role behavior in women and men with congenital adrenal hyperplasia (CAH). J. Sex Res. 41, 75–81.

Hooven, C.K., Chabris, C.F., Ellison, P.T., Kosslyn, S.M., 2004. The relationship of male testosterone to components of mental rotation. Neuropsychologia 42, 782–790.

Hugdahl, K., Thomsen, T., Ersland, L., 2006. Sex differences in visual–spatial processing: an fMRI study of mental rotation. Neuropsychologia 44, 1575–1583.

Jordan, K., Wustenberg, T., Heinze, H.-J., Peters, M., Jancke, L., 2002. Women and men exhibit different cortical activation patterns during mental rotation tasks. Neuropsychologia 40, 2397–2408.

Kempel, P., Gohlke, B., Klempau, J., Zinsberger, P., Reuter, M., Hennig, J., 2005. Second-to-fourth digit length, testosterone and spatial ability. Intelligence 33, 215–230.

Linn, M.C., Petersen, A.C., 1974. Emergence and characterization of sex differences in spatial ability: a meta-analysis. Child Dev. 56, 1479–1498.

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., Manning, J.T., 2004. 2nd to 4th digit ratios, fetal testosterone and estradiol. Early Hum. Dev. 77, 23–28.

Maccoby, E.E., Jacklin, C.N., 1974. The Psychology of Sex Differences. Stanford University Press, Stanford CA.

Manning, J.T., Taylor, R., 2002. Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans. Evol. Hum. Behav. 21, 61–69.

Manning, J.T., Scott, D., Wilson, J., Lewis-Jones, D.I., 1998. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. Hum. Reprod. 13, 3000–3004.

Manning, J.T., Bundred, P.E., Newton, D.J., Flanagan, B.F., 2003. The second to fourth digit ratio and variation in the androgen receptor gene. Evol. Hum. Behav. 24, 399–405.

Martin, C.L., 1999. A developmental perspective on gender effects and gender concepts. In: Swann Jr., W.B., Langlois, J.H. (Eds.), Sexism and Stereotypes in Modern Society: The Gender Science of Janet Taylor Spence. American Psychological Association, Washington, DC, USA, pp. 45–73.

McFadden, D., Shubel, E., 2003. The relationships between otoacoustic emissions and relative lengths of fingers and toes in humans. Horm. Behav. 43, 421–429.

McIntyre, M.H., 2006. The use of digit ratios as markers for perinatal androgen action. Reproductive Biology and Endocrinology 4, 10–18.

Merriman, W.E., Keating, D.P., List, J.A., 1985. Mental rotation of facial profiles: age-, sex-, and ability-related differences. Developmental Psychology 21, 888–900.

Miles, C., Green, R., Hines, M., 2006. Estrogen treatment effects on cognition, memory and mood in male-to-female transsexuals. Horm. Behav. 50, 708–717.

Packard, M.G., Teather, L.A., 1998. Amygdala modulation of multiple memory systems: hippocampus and caudate-putamen. Neurobiol. Learn. Mem. 69, 163–203.

Parsons, L.M., 2003. Superior parietal cortices and varieties of mental rotation. Trends Cogn. Sci. 7, 515–517.

Peters, M., Laing, B., Latham, D., Jackson, M., Zaiyouna, R., Richardson, C., 1995. A redrawn Vandenberg and Kuse mental rotations test: different versions and factors that affect performance. Brain Cogn. 28, 39–58.

Platek, S.M., Keenan, J.P., Mohamed, F.B., 2005. Sex differences in the neural correlates of child facial resemblance: an event-related fMRI study. Neuroimage 25, 1336–1344.

Poulin, M., O'Connell, R.L., Freeman, L.M., 2003. Picture recall skills correlate with 2D:4D ratio in women but not in men. Evol. Hum. Behav. 25, 174–181.

Putz, D.A., Gaulin, S.J.C., Sporter, R.J., McBurney, D.H., 2004. Sex hormones and finger length: what does 2D:4D indicate? Evol. Hum. Behav. 25, 182–199.

Rossman, B.B., Horn, J.L., 1972. Cognitive, motivational and temperamental indicants of creativity and intelligence. J. Educ. Meas. 9, 265–286.

Ruble, D.N., Martin, C.L., 1998. Gender development. In: Eisenberg, N. (Ed.), Handbook of Child Psychology. Wiley, New York, pp. 933–1016.

Shepard, R.N., Metzler, J., 1971. Mental rotation of three-dimensional objects. Science 171, 701–703.

Sherman, J., 1967. Problem of sex differences in space perception and aspects of intellectual functioning. Psychol. Rev. 290–299.

Tracy, D.M., 1987. Toys, spatial ability, and science and mathematics achievement: are they related? Sex Roles 17, 115–138.

Vandenberg, S.G., Kuse, A.R., 1978. Mental rotation, a group test of three-dimensional spatial visualisation. Percept. Mot. Skills 47, 599–604.

Voyer, D., 1997. Scoring procedure, performance factors, and magnitude of sex differences in spatial performance. Am. J. Psychol. 110, 259–276.

Voyer, D., Hou, J., 2006. Type of items and the magnitude of gender differences on the Mental Rotations Test. Can. J. Exp. Psychol. 60, 91–100.

Voyer, D., Voyer, S., Bryden, M.P., 1995. Magnitude of sex differences in spatial abilities: a metal-analysis and consideration of critical variables. Psychol. Bull. 117, 250–270.

Voyer, D., Nolan, C., Voyer, S., 2000. The relation between experience and spatial performance in men and women. Sex Roles 43, 891–915.

Yamazaki, K., Hirata, K., Mimuro, I., Kaitoh, Y., 2001. A case of dressing apraxia: contributory factor to dressing apraxia. J. Neurol. 248, 235–236.